Research article

Multi-task learning for aspect level semantic classification combining complex aspect target semantic enhancement and adaptive local focus

Quan Zhu[1,2,*], Xiaoyin Wang[2], Xuan Liu[1,2], Wanru Du[1,2] and Xingxing Ding[1,2]

[1] China Aerospace Academy of Systems Science and Engineering, Beijing 100048, China
[2] Aerospace Hongka Intelligent Technology (Beijing) CO., LTD., Beijing 100048, China

* Correspondence: Email: judisquan@163.com; Tel: +8618185301909.

Abstract: Aspect-based sentiment analysis (ABSA) is a fine-grained and diverse task in natural language processing. Existing deep learning models for ABSA face the challenge of balancing the demand for finer granularity in sentiment analysis with the scarcity of training corpora for such granularity. To address this issue, we propose an enhanced BERT-based model for multi-dimensional aspect target semantic learning. Our model leverages BERT's pre-training and fine-tuning mechanisms, enabling it to capture rich semantic feature parameters. In addition, we propose a complex semantic enhancement mechanism for aspect targets to enrich and optimize fine-grained training corpora. Third, we combine the aspect recognition enhancement mechanism with a CRF model to achieve more robust and accurate entity recognition for aspect targets. Furthermore, we propose an adaptive local attention mechanism learning model to focus on sentiment elements around rich aspect target semantics. Finally, to address the varying contributions of each task in the joint training mechanism, we carefully optimize this training approach, allowing for a mutually beneficial training of multiple tasks. Experimental results on four Chinese and five English datasets demonstrate that our proposed mechanisms and methods effectively improve ABSA models, surpassing some of the latest models in multi-task and single-task scenarios.

Keywords: aspect-based sentiment classification; BERT language model; adaptive local attention mechanism; multiple-task learning

1. Introduction

Aspect-based sentiment analysis (ABSA) [1–5] is a task that aims to uncover the most fine-grained aspects of sentiment information in the text. The main goal of this task is to identify the existing aspect targets in the corpus and their corresponding sentiment polarity.

According to existing models and composite sentiment element extraction methods, ABSA can be further divided into aspect-opinion pair extraction (AOPE), aspect category sentiment analysis (ACSA) and aspect target sentiment classification (ATSC) [6]. AOPE extracts aspect and opinion terms in pairs to clearly describe the aspect target and the corresponding opinion expression. ACSA jointly detects the category to which the discussed aspect belongs and its corresponding sentiment polarity. ATSC includes two subtasks, aspect term extraction (ATE) and aspect sentiment classification (ASC), which extract sentiment information pairs (aspect target, polarity) from the text. This paper mainly focuses on the deep learning-based ATSC task.

In recent years, ATSC has received increasing attention due to its widespread application in sentiment analysis. Neural networks in NLP, such as long short-term memory networks (LSTM) [7–11], convolutional neural networks (CNN) [12] and memory networks (MemNet) [13,14] have been widely used in sentiment classification tasks, while the emergence of Transformer [15] has successfully achieved end-to-end training. However, compared to document-level and sentence-level sentiment analysis tasks [16,17], ATSC tasks require distinguishing the sentiment of different aspect targets, which requires more subtle semantic features for model analysis [18]. Therefore, there are still some challenges and many unresolved issues: 1) Fully supervised tasks require a large amount of text data manually labeled by humans, and commonly used corpora are small, making it difficult for models to learn and identify complex sentences containing different aspect targets of sentiment elements. 2) Sentences of different lengths, due to the different positions of aspect entities in the sentences, will produce varying degrees of noise information that can reduce the model's accuracy in recognizing sentiment information. 3) Training for ATE and ASC tasks alone does not consider the need for entirely accurate sentiment information pairs *(aspect target, polarity)*, and the design of single tasks can reduce the model's efficiency during task completion.

In this paper, we aim to address the issues mentioned above by improving and combining existing semantic understanding models, leading to the proposal of an adaptive sentiment semantic model that targets the complex semantic enhancement of aspect targets. The model can expand different aspect targets information in multiple corpora, which enriches the training sample set and expands the semantic understanding scope of the model. Second, to highlight the importance of the aspect target's semantic features to the model's learning, we propose the aspect recognition enhancement mechanism. Then, we hope to automatically adjust the semantic understanding scope of the text centered on the aspect target according to the expanded different corpus so that the model can recognize more specific emotional polarity views corresponding to the aspect target. Finally, the improvement enables the model to achieve a stable multi-task joint learning approach by considering ATE task and ASC task simultaneously, which facilitates the accurate recognition of sentiment information pairs (aspect target, polarity). Our model rigorously ensures that the augmented dataset remains from the same domain, accomplishing the ability to flexibly adjust the semantic embeddings of aspect targets within the same sentence based on the aspect target to be evaluated. This adjustment makes the embeddings more inclined to represent the sentiment elements associated with the aspect target, thereby achieving a finer-grained level of sentiment analysis.

The contributions of this paper can be summarized as follows:

We propose a complex semantic enhancement model based on BERT [19] (BERT-CSE), which

improves upon SBERT [20] and enables it to adapt perfectly to sentiment semantic understanding in ABSA. The model highlights the aspect target as the center of attention in the attention mechanism for semantic enhancement, enriching the small-scale sentiment corpus.

1) We propose an aspect recognition enhancement mechanism to semantically fuse global contextual semantic features and local contextual semantic features with the aspect target semantic features, which enriches the overall semantic features.

2) We propose an adaptive global-local attention mechanism sentiment recognition model based on BERT (BERT-ASD), which can limit the effective local text length range for different texts, thus reducing the negative impact of noisy sentiment elements brought to the model by redundant text data.

3) To achieve the ATSC task more efficiently, we implement and improve the joint learning mechanism so that the model can learn more stably and effectively on the ATE task and the ASC task simultaneously.

The subsequent structure of the paper is as follows: Section 2 provides a summary and detailed introduction of related work. In Section 3, we first define the ATSC task and then describe the mechanisms and models proposed in our work. Section 4 describes the datasets and experimental settings used for evaluation, presents the experimental results and discusses the overall experimental results. Finally, Section 5 summarizes our work.

2. Related works

2.1. Machine learning-based methods

In general, for traditional classification tasks, machine learning methods include decision trees [21], KNN [22], Naive Bayes [23], logistic regression [24], support vector machines (SVM) [25], random forests [26] and so on. These methods can also be used for other subtasks, each with unique advantages. The most suitable machine learning algorithm for sentiment classification tasks can be selected based on the data distribution. Although SVM performs best in document-level sentiment classification tasks [27], its performance still needs to be improved for ABSA. Additionally, traditional machine learning methods rely heavily on a large number of well-designed manual feature engineering to a certain extent and feature engineering is already a tedious task that increases its manual and time costs in the era of big data.

2.2. Deep learning-based methods

Deep learning methods use neural networks with complex parameter structures. Compared to machine learning methods, they save a lot of feature engineering work and have better results in aggregating data information. At the initial stage, neural networks had limited data samples and were restricted by hardware limitations, but with the advent of the big data era and the improvement of industrial technology, deep learning methods have flourished and have been applied to many other tasks in NLP, such as text generation, machine translation, question-answering systems, entity relationship extraction, etc. [28]. Due to the many variants of neural networks, they can flexibly handle various forms of data and aggregate rich semantic and emotional information for ABSA tasks [29–32]. In the following, we will discuss several commonly used neural network models for sentiment analysis.

2.2.1. Methods based on recurrent neural network

Xu et al. [33] utilized the position attention mechanism for weighting the output of LSTM, enhancing the feature representation ability of the model and expanding the knowledge base information. Ma et al. [34] combined a commonsense knowledge embedding layer with an attention mechanism. The commonsense knowledge embedding layer incorporated domain-specific commonsense understanding into the model, while the LSTM with attention mechanism further enhanced the model's generalization ability. To address the problem of word normalization, Bao et al. [35] employed lexical normalization techniques and used two LSTM layers, one for learning sentence-level sentiment features and the other for learning aspect-level sentiment features.

2.2.2. Methods based on convolutional neural network

Xing et al. [36] was the first study to apply CNN to aspect-level sentiment classification, demonstrating its effectiveness in this task. Wang et al. [37] proposed a novel PCNN model that divides the text into multiple positions, enabling the integration of positional information with the convolutional neural network and better helping the model to understand the importance of different positions in the text. Gan et al. [38] proposed SA-SDCNN, which combines sparse attention mechanism, separable convolution and dilated convolution to improve the performance of targeted sentiment analysis without using pre-trained word embeddings. Zhao et al. [39] proposed CR-CNN, which extracts features of each word in the text using CNN and learns the dependencies between the features using the GRU model. They also introduced a gated mechanism to help the model better understand the complex relationship between aspect and sentiment words.

2.2.3. Methods based on memory network

Yi et al. [40] proposed a model based on dyadic memory networks. They used a convolutional neural network-based bidirectional encoder to encode the input sentence and a dyadic memory network based on LSTM to capture the relationships between aspect words. Chen et al. [41] proposed HMAN, which includes a multi-head self-attention layer and a multi-head interaction layer for encoding and interacting with the text sequence and aspect target. Zhang et al. [42] proposed a memory-based convolutional multi-head self-attention model that uses a memory network to encode previous information for retrieval of important information during classification. The relationship between aspect target, sentiment and context can be well captured.

2.2.4. Methods based on transformer

Song et al. [43] proposed AEN, a model consisting of a target word extractor and a sentiment classifier. The target word extractor uses a gating mechanism to extract the target word from the sentence and employs a convolutional neural network for feature extraction. The sentiment classifier is composed of an LSTM network based on an attention mechanism. Yang et al. [44] proposed a local attention mechanism model to implement a multi-task learning mechanism with good performance on Chinese datasets. Akbar et al. [45] introduced a hierarchical summarization mechanism and sentiment lexicon information to expand the original unidirectional summarization mechanism into a bidirectional one, which is used for adjusting the output of the BERT model. Akbar et al. [46] proposed an adversarial training method based on BERT (ABSA-AT), consisting of adversarial samples and

training. Adversarial sample generation creates a set of adversarial samples by making small perturbations to the original input text. Adversarial training uses these adversarial samples along with the original samples for model training.

3. Methods

3.1. Task definition

The purpose of aspect-based sentiment classification (ATSC) is to extract the sentiment polarity of the comment information corresponding to an aspect target as the center from a sentence, forming the correct sentence sentiment information pair $(aspect\ target, polarity)$. This task can be further divided into named entity recognition (NER) and classification tasks. Given a corpus set S=$\{s_1, s_2, s_3, \ldots, s_n\}$ and its corresponding aspect target set E=$\{e_1, e_2, e_3, \ldots, e_n\}$, the sentiment polarity of the sentence is extracted with respect to the aspect target, forming a correct sentiment aspect pair (aspect target, polarity), which can be divided into NER and classification tasks. The sentiment classification set is $C = positive, neutral, negative$, and the label set for entities is $L = B, I, O, [CLS], [SEP], 0$. The proposed aspect target complex semantic-enhanced model based on BERT is denoted as $M_1(\bullet)$, and the proposed adaptive global-local attention mechanism sentiment recognition model based on BERT is denoted as $M_2(\bullet)$. According to the two different tasks, they can be defined as follows:

$$c_i^* = \underset{i \in (1,n)}{argmax}\ P(c_i | M_2(M_1(s_i), e_i)) \tag{1}$$

$$l_i^* = \underset{i \in (1,n)}{argmax}\ P(l_i | M_2(M_1(s_i), e_i)) \tag{2}$$

where $P(\bullet)$ denotes the predicted probability distribution, c_i^* denotes the set of true sentiments corresponding to e_i in sentence s_i and l_i^* denotes the true labels corresponding to each e_i in sentence s_i.

3.2. BERT-based aspect target complex semantic enhancement model

In response to the problem of short texts and limited aspect targets in small sentiment analysis corpora, which are insufficient to support the identification of more complex and fine-grained emotional information, we propose a data augmentation model centered on aspect targets. This model is based on SBERT and uses triplet networks to build its framework, as shown in the model structure diagram. The construction of the model mainly depends on the training data format we want to enhance. We plan to combine the corpus sentences according to the model's learning pattern to expand the training sample set with complex emotional semantics. Figure 1 shows the overall structure of our model.

3.2.1. Self-supervised processing of complex semantics

Since the data format of the corpus does not match the input format of the pre-trained model BERT, data pre-processing is required before feeding it to the model. The corpus contains many utterances, and manual labeling reduces the overall efficiency of the task solution. Therefore, we

propose a data augmentation model that focuses on the aspect target in the sentence. We randomly sample the corpus S_m to collect m data pairs $(s_i, e_i, s_j, category)$ with $category = \{0, 1\}$, where s_i and s_j are two random sentences, and e_i corresponds to the aspect target of s_i. We first iterate through each sentence s_i in S_m, select another sentence s_j in S_m by random sampling, and determine the value of the category based on the word similarity Sim_{ij} of e_i corresponding to s_i and e_i corresponding to s_j. Finally, we add the extracted $(s_i, e_i, s_j, category)$ pairs to the set RES. The specific operational process is shown in Table 1.
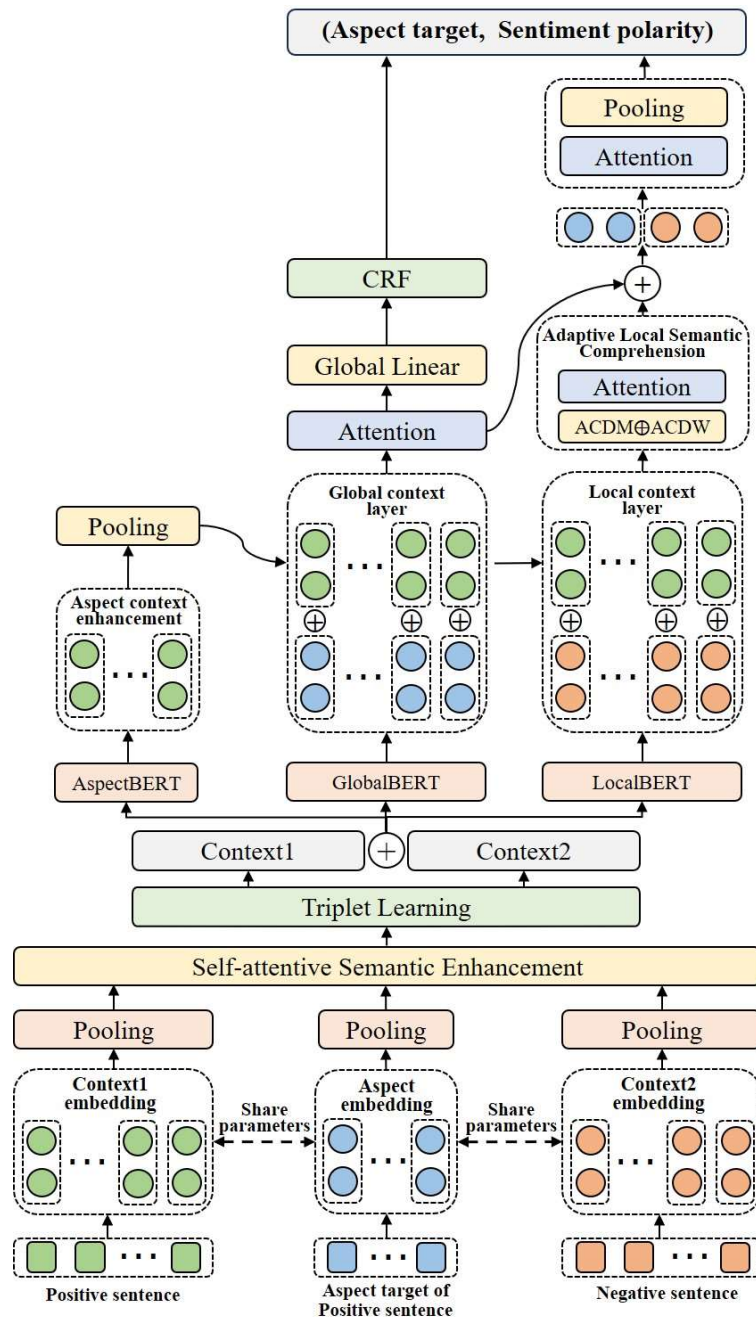

Figure 1. The framework structure of our BERT-ATSE model.

Table 1. RES construction process.

Collection Process
Input： corpus $S = \{s_1, s_2, s_3, \ldots, s_m\}$， aspect target $E = \{e_1, e_2, e_3, \ldots, e_m\}$
Output： Data pairs RES= $\{(s_i, e_i, s_j, category) \| i \leq m, j \leq m, j \neq i, category = \{0,1\}\}$
RES← $\emptyset$
for each $s_i \in S$ do
$\quad s_j \leftarrow random\_select(S)$
$\quad$ category← 0
$\quad$ while $s_i == s_j$ do
$\qquad s_j \leftarrow random\_select(S)$
$\quad$ end while
$\quad$ if $Sim_{ij} > 0.5$ do
$\qquad$ category← 1
$\quad$ end if
$\quad$ RES←$(s_i, e_i, s_j, category)$
end for

3.2.2. Embedded learning of semantic blocks

The processed data pairs $RES(s_i, e_i, s_j, category)$ are obtained, where s_i serves as the positive sample, e_i as the anchor sample, s_j as the negative sample and category is used to indicate whether the aspect target in the positive and negative samples is similar or dissimilar, with values of 0 and 1, respectively. The sentences s_i and s_j are concatenated as "[CLS] + Context+ [SEP] + aspect target + [SEP]" and e_i is attached as "[CLS] + aspect target + [SEP]". They are then converted into token sequences and indexed according to the BERT vocabulary, along with additional information such as position end segment embeddings. These token indices are input into BERT to obtain the context embedding sequences $h_b^p = \{h_1^p, h_2^p, \ldots, h_n^p\}$ for the positive sample's context, $h_b^a = \{h_1^a, h_2^a, \ldots, h_n^a\}$ for the aspect target's context and $h_b^n = \{h_1^n, h_2^n, \ldots, h_n^n\}$ for the negative sample's context.

$$S_{bert} = BERT(S) \tag{3}$$

3.2.3. Self-attentive semantic enhancement of aspect target

To focus more on the text information containing aspect targets for both positive and negative samples and reduce the negative impact caused by long-distance contextual dependencies on aspect target understanding, we input the context embeddings generated by the semantic understanding layer into a multi-head self-attention layer (MHSA). Assuming that $S = s_1, s_2, \ldots, s_n$ is the context embedding sequence input, W_q, W_k and W_v are three weight parameter matrices used to perform element-wise multiplication with S, mapping to three matrices Q, K and V, respectively. The multi-head self-attention layer (MHSA) is applied using the resulting matrices, and a new sequence $S^* = \{s_1^*, s_2^*, \ldots, s_n^*\}$ of the same length as the original sequence is obtained. The specific calculation process is shown as follows:

$$Self\,Attention(S) = SA(S, S) \tag{4}$$

$$SA(S, S) = V \cdot Softmax(\frac{K^T \cdot Q}{\sqrt{d^q}}), \tag{5}$$

$$\begin{cases} Q = W_q \cdot S \\ K = W_k \cdot S \\ V = W_v \cdot S \end{cases} \tag{6}$$

where $W_q \in \mathbb{R}^{d^q \times d^q}$, $W_k \in \mathbb{R}^{d^k \times d^k}$, $W_v \in \mathbb{R}^{d^v \times d^v}$, $S \in \mathbb{R}^{d^q \times d^s}$, $Q \in \mathbb{R}^{d^q \times d^s}$, $K \in \mathbb{R}^{d^k \times d^s}$ and $V \in \mathbb{R}^{d^v \times d^s}$. Here, $SA(\cdot)$ denotes the use of the self-attention mechanism, and $d^q = d^k = d^v$.

The above is the operation for only one head, while in MHSA, the output matrix needs to be transformed by the weight matrix $W_s \in \mathbb{R}^{d^q \times m \cdot d^q}$, and the final result is output through the tanh activation function.

$$MHSA(S) = tanh(W_s \cdot \{h_1; h_2; \dots; h_m\}) \tag{7}$$

where ";" denotes the vertical concatenation of vectors, m denotes the number of heads, $MHSA(\cdot)$ denotes the use of multi-head self-attention mechanism and W of the above equations are all parameters that need to be learned by the above equations.

$$S_{MHSA} = \text{MHSA}(S_{bert}) \tag{8}$$

$$S_{POOL} = POOLING(S_{MHSA}) \tag{9}$$

In the POOLING layer, three methods were selected during the experiment to learn sentence embeddings h'^p_b, h'^a_b and h'^n_b that are more capable of distinguishing between sentence semantics: using the vector of the [CLS] token to represent the overall context embedding ($CLSPooling$); calculating the average value of the vectors of each token in the sentence to represent the overall context embedding ($MeanPooling$); finding the maximum value vector by taking the maximum value of the word vectors in the sentence to represent the overall context embedding ($MaxPooling$).

Finally, the dimensionality of the vectors is transformed by a fine-tuned linear layer to transfer them to a dimension more suitable for sentence representation, and all fine-tuned sentence vectors are cached as set H_o and their corresponding sentences are cached as set S_o to enable the model to learn the semantic variability between sentences.

$$H_o = W_l \cdot S_{POOL} + b_l \tag{10}$$

where $W_{l1} \in \mathbb{R}^{d^{q1} \times d^q}$ is the weight representation and $b_{l1} \in \mathbb{R}^{d^{q1} \times d^s}$ is the bias representation.

3.2.4. Model learning

We employ triplet loss to train our model, in which each sentence has an opportunity to be a positive sample s_i, the corresponding anchor sample e_i is an aspect target in the positive sample and the negative sample comes from other sentences s_j in the random sample. We aim to use self-supervised data construction to facilitate the model in distinguishing subtle differences in sentence semantics. This is achieved by minimizing the following loss function:

$$Loss = max(||e_i, s_i||^2 - ||e_i, s_j||^2 + \varphi, 0) \tag{11}$$

Triplet loss is available in three cases:

a) Easy triplets, where $||e_i, s_i||^2 + \varphi < ||e_i, s_j||^2$, $Loss = 0$ and the corresponding data cannot make any suggestions for improving the parameters of the model.

b) Semi-hard triplets, where $||e_i, s_i||^2 < ||e_i, s_j||^2 < ||e_i, s_i||^2 + \varphi$, $0 < Loss < \varphi$. At this point, the model can slightly understand the positive and negative samples in the data pairs but cannot make a clear distinction.

c) Hard triplets, where $||e_i, s_i||^2 > ||e_i, s_j||^2$, $Loss > \varphi$. The model incorrectly distinguishes between positive and negative samples in the data pairs and is able to improve the transfer parameters of the model the fastest.

3.2.5. Unsupervised construction of complex semantic samples

Affected by the short text of the corpus, it may not be possible for downstream sentiment recognition models to accurately identify every aspect target and its corresponding sentiment polarity in long texts. Therefore, we need to use this model to improve and enhance the data initially. We aim to concatenate the two sentences s_i and s_j with the smallest semantic difference as a training data point $(s_i; s_j)$. If the downstream model can make accurate judgments on the $(aspect, polarity)$ information presented in the low semantic difference between the sentences $Sim(s_i; s_j)_{low}$, then we assume that it can also make correct decisions for sentences $Sim(s_i; s_j)_{high}$ with high semantic differences.

To better represent the similarity between two sentences, we use Manhattan distance to calculate $Sim(s_i; s_j)$, where x and y represent the sentence vectors corresponding to s_i and s_j, respectively. The symbol $|\cdot|$ denotes the absolute value, and x_i and y_i represent the i-th element in the vectors:

$$Sim(s_i; s_j) = |x_1 - y_1| + |x_2 - y_2| + \cdots + |x_n - y_n| \tag{12}$$

We randomly select a sentence s_j from the cached S_o set for each sentence s_i, with a similarity score greater than the similarity threshold ssh. We then combine the sentences s_i and s_j to create a new complex semantic sample dataset, denoted as S'.

3.3. Adaptive global-local attention mechanism for sentiment analysis based on BERT

This section proposes a BERT-based adaptive global-local attention mechanism for joint learning of aspect targets and their corresponding sentiment polarities. Here is the architecture of our model, which employs the following mechanisms:

1) BERT-based semantic understanding learning: the model includes 2 BERT models. One BERT is used to learn the semantic features of aspect target words and global sentence features, while the other BERT is used to understand the local sentence semantic features.

2) Semantic enhancement with aspect target as the core: We perform average pooling on the semantic features of aspect entities in the sentence and concatenate them with global and local semantic features. Then, through a linear layer and an MHSA layer, we input them into the adaptive local

semantic hidden layer and global semantic hidden layer, enhancing the core semantic features of the aspect target.

3) Adaptive local semantic understanding: In our proposed BERT-CSE, the lengths of sentences in the updated text corpus vary. The adaptive local semantic understanding enables the model to focus on the most effective sentiment information near the aspect target, eliminating redundant sentiment information and ensuring that the model is not affected by the length of the text data.

4) Adopting synchronous joint learning mechanism: We bind the ATE and ASC tasks and improve the backpropagation of the overall loss to achieve effective and stable joint learning of multiple tasks.

3.3.1. Embedded learning of semantics

We employ BERT as the basic semantic learning architecture and learn the global semantic feature by constructing the sequence of "[CLS] + context + [SEP]", which is then fed into the BERT model. Let the hidden layer dimension of each token in the BERT model be d_b, and the number of tokens in each sentence be m_1. $GlobalBERT(\cdot)$ denotes the use of the hidden layer parameters of the BERT model for learning global semantics. For the new dataset $S' = \{s'_1, s'_2, ..., s'_n\}$ we constructed, the global sequence is constructed, and the dimension of the input sample after being fed into the BERT model is S^g, with each s'_{gi},

$$S^g = GlobalBERT(S') = \{s'_{g1}, s'_{g2}, ..., s'_{gn}\} \tag{13}$$

where $S^g \in \mathbb{R}^{n \times m_1 \times d_b}$ and $s'_{gi} \in \mathbb{R}^{m_1 \times d_b}$.

For the learning of local semantic features, we aim to focus on the aspect target and deepen the model's understanding of it. Here, we use the BERT-SPC input mode and concatenate the sentence as "[CLS] + context + [SEP] + aspect target + [SEP]", which is input to the BERT model in the same way. $LocalBERT(\cdot)$ denotes the hidden layer parameters of the learned local semantic BERT model. The sequence obtained by entering the input into BERT is S^l, with each s'_{li}.

$$S^l = LocalBERT(S') = \{s'_{l1}, s'_{l2}, ..., s'_{ln}\} \tag{14}$$

where $S^l \in \mathbb{R}^{n \times m_1 \times d_b}$ and $s'_{li} \in \mathbb{R}^{m_1 \times d_b}$.

The semantic representation of each e'_i in the new set $E' = \{e'_1, e'_2, ..., e'_n\}$ of aspect targets is learned by $GlobalBERT(\cdot)$, and m_2 denotes the number of tokens of each sentence aspect target, where $AspectBERT(\cdot)$ denotes the hidden layer parameters of the semantic BERT model using the learned aspect target, and the sequence E^b obtained by entering BERT.

$$E^b = AspectBERT(E') = \{e'_{b1}, e'_{b2}, ..., e'_{bn}\} \tag{15}$$

$$E^p = AVGPOOLING(E^b) \tag{16}$$

We apply an average pooling layer to construct the average semantic vector set E^p for the aspect targets. $AVGPOOLING(\cdot)$ represents the hidden layer parameters of the average pooling layer. Where $E^b \in \mathbb{R}^{n \times m_2 \times d_b}$ and $E^p \in \mathbb{R}^{n \times 1 \times d_b}$.

3.3.2. Aspect recognition enhancement mechanism

The subsequent work of identifying the sentiment polarity is valuable only by accurately identifying the aspect target in the sentence. To further emphasize the importance of aspect targets in learning the interaction of text sentiment elements, we perform semantic feature interaction learning between aspect targets, global semantics and local semantics.

First, E^p is concatenated with S^l and S^g in the second dimension, respectively. Then, their information is compressed and fused using the linear layer. The compressed and enhanced information is then aggregated through a multi-head self-attention layer. As an example of global semantic feature interaction learning, the semantic enhancement process with the aspect target as the core is calculated as follows, resulting in the semantic enhancement effect of the aspect target:

$$S^{ge} = [S^g; E^p] = \{s'_{g1}; e'_{p1}, s'_{g2}; e'_{p2}, \dots, s'_{gn}; e'_{pn}\} \tag{17}$$

$$S^{ge}_{dense} = S^{ge} \cdot W^{ge} + b^{ge} \tag{18}$$

$$S^{ge}_{MHSA} = MHSA(S^{ge}_{dense}), \tag{19}$$

where $W^{ge} \in \mathbb{R}^{2 \cdot d_b \times d_b}$, $b^{ge} \in \mathbb{R}^{m_1 \times d_b}$ are the parameters of the linear layer, the $S^{ge} \in \mathbb{R}^{n \times m_1 \times 2 \cdot d_b}$ is obtained after concatenation and S^{ge}_{dense}, $S^{ge}_{MHSA} \in \mathbb{R}^{n \times m_1 \times d_b}$.

3.3.3. Extraction of aspect targets

We define a label set L for aspect-based sentiment analysis, where B represents the initial token of the aspect target, I represents the internal and tail tokens of the aspect target, O indicates other tokens in the sentence except for the aspect target and $[CLS]$ and $[SEP]$ are respectively the beginning and separation representations in the BERT input mode. 0 denotes the padding part of the sentence. We aim to use the vector S^{ge}_{MHSA}, which is obtained through global semantic comprehension, to complete the ATE task. First, the dimensions are transformed into six dimensions of the set L by linear parameters $W^{t_1} \in \mathbb{R}^{d_b \times 6}$, $b^{t_1} \in \mathbb{R}^{m_1 \times 6}$:

$$S^{ge}_{t_1} = S^{ge}_{MHSA} \cdot W^{t_1} + b^{t_1}. \tag{20}$$

Given a subsequence of sentences $x = \{w_1, w_2, \dots, w_n\}$ and a sequence of predicted labels $y = \{y_1, y_2, \dots, y_n\}$, the scores of each label of the predicted sequence are obtained when passing through the CRF model:

$$score(x, y) = \sum_{i=1}^{n}(p_{i,y_i} + t_{y_{i-1}, y_i}), \tag{21}$$

where p_{i,y_i} denotes the probability score of the i-th token labeled y_i and t_{y_{i-1}, y_i} denotes the probability score of the label y_{i-1} transferred to y_i, and then the sequence is normalized to obtain the prediction sequence.

$$P(y|x) = \frac{e^{score(x,y)}}{\sum_{y' \in y_x} e^{score(x,y')}} \tag{22}$$

where y' denotes the true label value and y_x denotes the set of all possible labels. The final dynamic programming algorithm of the Viterbi algorithm is used to solve the optimal entity recognition sequence with probability scores $y^e \in \mathbb{R}^{1 \times n}$:

$$y^e = \underset{y' \in y_x}{argmax}\, score(x, y) \tag{23}$$

3.3.4.   Adaptive local semantic comprehension

Adaptive local semantic understanding refers to mechanisms that can adaptively focus on understanding local sentiment semantics depending on the text of different lengths. Semantic relative distance (SRD) [44] has been proven to be effectively utilized for focusing on local contextual semantics, and SRD is calculated as follows:

$$SRD_i = |i - p_a| - \left\lfloor \frac{m}{2} \right\rfloor \tag{24}$$

where SRD_i denotes the SRD value of the ith token in the sentence, i denotes the index of the token's position in the sentence, p_a denotes the index of the center position of the aspect target and m denotes the length of the aspect target.

We aim to focus our attention on the vicinity of the aspect target. Based on data analysis, we have designed an adaptive local range function to determine the focus range of the model so that each sentence has its appropriate semantic threshold. l_a represents the number of words in the aspect target, l_t represents the number of words in the sentence and i_m represents the index position of the center word in the aspect target within the sentence.

$$\alpha_w = \log_{l_a+2} l_t + \log_{l_a+1} i_m - \frac{l_a}{3} \tag{25}$$

Next, we incorporate this function into the adaptive context dynamic mask (ACDM) and the adaptive context dynamic weighted (ACDW). ACDM is used to better reduce the interference of noisy semantics for data with a longer threshold range. On the other hand, ACDW is utilized to balance the elimination of redundant semantics and the preservation of the integrity of the main semantics for data with a shorter threshold range.

ACDM sets an initial matrix T_M with values of 0 to represent the mask of a sentence. Each t_i^M is computed by taking the dot product of a vector O consisting of all 0 and a vector I consisting of all 1, as shown in the equation below. The resulting matrix is denoted as S_{ACDM}^{le}.

$$t_i^M = \begin{cases} O, SRD_i > \alpha_w \\ I, SRD_i \le \alpha_w \end{cases} \tag{26}$$

$$T_M = \left[t_1^M, t_2^M, \dots, t_{m_1}^M \right] \tag{27}$$

$$S_{ACDM}^{le} = S_{MHSA}^{le} \cdot T_M \tag{28}$$

where matrix $T_M \in \mathbb{R}^{m_1 \times d_b}$ with each t_i^M, O, $I \in \mathbb{R}^{1 \times d_b}$ and $S_{ACDM}^{le} \in \mathbb{R}^{n \times m_1 \times d_b}$ is the result.

ACDW adopts a scheme that decreases the weights according to SRD in a hierarchical manner. We set a vector matrix T_W to represent each t_i^W, which needs to be calculated using the formula, where $\cdot$ denotes element-wise multiplication. The resulting matrix is denoted as S_{ACDW}^{le}.

$$t_i^M = \begin{cases} \left(1 - \frac{SRD_i - \alpha_w}{m_1}\right) \cdot I, SRD_i > \alpha_w \\ I, SRD_i \le \alpha_w \end{cases} \tag{29}$$

$$T_W = \left[t_1^W, t_2^W, \dots, t_{m_1}^W\right] \tag{30}$$

$$S_{ACDW}^{le} = S_{MHSA}^{le} \cdot T_W \tag{31}$$

where matrix $T_W \in \mathbb{R}^{m_1 \times d_b}$ with each $t_i^W \in \mathbb{R}^{1 \times d_b}$, and $S_{ACDW}^{le} \in \mathbb{R}^{n \times m_1 \times d_b}$. These two masking approaches help to alleviate the interference of noisy sentiment information of varying lengths on the model after we construct the new data set S'.

3.3.5. Global-local semantic feature interaction learning

To enrich the semantic feature information required by the model, it continues to fuse the features of global-local semantics by concatenating S_{MHSA}^{ge} with S_{ACDM}^{le} and S_{ACDW}^{le} to obtain S_{whole}^{gle}. The hidden layer parameters W^{t_2} and b^{t_2} are used to obtain $S_{t_2}^{gle}$, and then the global-local semantics are aggregated by MHSA to obtain S_{MHSA}^{gle}.

$$S_{whole}^{gle} = \left[S_{MHSA}^{ge}; S_{ACDM}^{le}; S_{ACDW}^{le}\right] \tag{32}$$

$$S_{t_2}^{gle} = S_{whole}^{gle} \cdot W^{t_2} + b^{t_2} \tag{33}$$

$$S_{MHSA}^{gle} = MHSA\left(S_{t_2}^{gle}\right) \tag{34}$$

where $S_{MHSA}^{ge}, S_{ACDM}^{le}, S_{ACDW}^{le} \in \mathbb{R}^{n \times m_1 \times d_b}$, $S_{whole}^{gle} \in \mathbb{R}^{n \times m_1 \times 2 \cdot d_b}$ and $W^{t_2}, b^{t_2} \in \mathbb{R}^{2 \cdot d_b \times d_b}$, $S_{t_2}^{gle} \in \mathbb{R}^{n \times m_1 \times d_b}$ and $S_{MHSA}^{gle} \in \mathbb{R}^{n \times m_1 \times d_b}$ are the results.

3.3.6. Classification of aspect polarity

The average pooling layer is used to obtain average vectors of the sequence vectors, and then the $softmax$ function is used to activate S_{MHSA}^{gle}, S_x^{gle} is a set of vectors containing information about all sentiment polarities, and $p\left(y|S_x^{gle}\right)$ denotes the predicted probability of obtaining a sentiment polarity given a sentiment representation S_x^{gle}.

$$S_{pool}^{gle} = AVGPOOL(S_{MHSA}^{gle}), \tag{35}$$

$$S_x^{gle} = S_{pool}^{gle} \cdot W^p + b^p \tag{36}$$

$$p\left(y|S_x^{gle}\right) = \frac{\exp(S_x^{gle})}{\sum_{i=1}^{d_y} \exp(S_x^{gle})} \tag{37}$$

where $S_{pool}^{gle} \in \mathbb{R}^{n \times d_b}$, W^p, $b^p \in \mathbb{R}^{d_b \times d_y}$, $S_x^{gle} \in \mathbb{R}^{n \times d_y}$ and d_y is the number of categories of sentiment polarity.

3.3.7. Model learning

1) ATE task loss processing

The training process of the ATE task is determined by the loss of the CRF model:

$$loss_{ate} = \ln\big(P(y|x)\big) = \ score(x,y) - \ln(\textstyle\sum_{y' \in y_x} e^{score(x,y')}). \tag{38}$$

2) ASC task loss processing

In the ASC task, we adopt a cross-entropy loss function, where $\hat{y}_i$ is the predicted value, y_i is the true value, C denotes the number of types of sentiment polarity, λ is a hyperparameter of L_2 regularization and Θ denotes all parameters used in the sentiment polarity classification task:

$$loss_{asc} = \sum_{i=1}^{C} \hat{y}_i \log y_i + \lambda \sum_{\theta_p \in \Theta} \left\| \theta_p \right\|_2 \tag{39}$$

3) Overall model loss processing

The model is designed to handle multiple tasks, but balancing the learning speed and parameter magnitude of these two tasks is difficult. To balance the contribution of the loss values of the two tasks during training, we dynamically calculate the average loss value of the two tasks for each batch and then use a learnable parameter α for dynamic weighted averaging. α needs to be mapped to the range of 0 to 1 through the $sigmoid$ function in each batch.

$$loss(\theta_{ate}, \theta_{asc}) = \alpha * loss_{ate} + (1 - \alpha) * loss_{asc} \tag{40}$$

During prediction, the correct sentiment information pairs $(aspect, polarity)$ are what we expect to obtain as results. The following represents the overall task studied in this paper, where l_i^* is the predicted entity recognition sequence and c_i^* is the predicted sentiment polarity classification sequence:

$$(l_i^*, c_i^*) = \underset{i \in (1,n)}{argmax}\, P((l_i, c_i) | M_2(M_1(s_i), e_i). \tag{41}$$

4. Experiments

4.1. Datasets

To comprehensively evaluate the performance of our proposed model, we used nine widely-used benchmark datasets. Four concise and diverse Chinese datasets from the COAE-2008 corpus were used for the model evaluation. Following previous works [47,48], we performed cleaning and rearranging on the original datasets, which have a sentiment distribution of positive and negative for each aspect target. Similarly, we used 4 SemEval datasets: SemEval-2014 task 4 [3], SemEval-2015 task 12 [4], SemEval-2016 task 5 [5] and Twitter for the ACL14 task [49]. The sentiment polarity of these datasets is negative, neutral and positive. Following previous works [50,51], we removed several data points with conflicting sentiment classification and aspect term extraction annotations. Almost all

of these datasets exhibit imbalanced sentiment distributions: Table 2 shows that most samples in the 4 Chinese datasets are positive, with negative samples accounting for only about half of the positive ones. Among them, Twitter has the most samples, with more subtle emotional elements and a higher proportion of neutral emotions, which poses a challenge to the model's sentiment recognition and reflects the practicality of our model.

4.2. Experimental settings

To train the Chinese and English datasets separately, we use bert-base-Chinese and bert-base-uncased applied to the model, which both have 12 transformer layers, 768 hidden layer units, 110M counts of parameters and a dropout rate of 0.1. In the training process of our model, Adam optimizer [52] is used for training with learning rate of 2×10^{-5}, batch size of 16, default training epochs of 10, the semantic similarity threshold ssh of 0.6 and the initial value for dynamic weighting α of the loss function is set to 0.5.

We employ a normal distribution to generate random numbers for initializing the weights and biases of the neural network. The generated random numbers are assigned to the respective parameter variables, serving as the initial parameters of the network. The hyperparameter settings of the model follow previous parameter setting experience [43,44]. As for the thresholds used in the model, they are initially randomly initialized within reasonable ranges. Subsequently, through continuous training experiments, the optimal threshold settings are determined by comparing the experimental results. All the training hyperparameters and thresholds are set as shown in Table 3.

During the training process on different datasets, we conducted controlled experiments to adjust hyperparameters for optimal results. Therefore, there may be some changes in hyperparameters, which will be discussed in detail in Section 4.7.

We report the Macro-average F1 (M-F1) for aspect target extraction task and the accuracy (Acc) and M-F1 values for the sentiment polarity classification task on these nine datasets. Tables 3 and 4, respectively, show the performance of various baseline models and advanced models on Chinese and English datasets for aspect target extraction and sentiment polarity classification tasks, demonstrating the potential of our model in multilingual tasks. Table 5 shows the overall performance of our model and other models on the ASC task in 5 English datasets. Table 6 shows the results of our model's ablation experiments in the BERT-BASE environment.

Table 2. Nine English and Chinese datasets used for evaluation.

Datasets	Negative		Neural		Positive		Total	
	Train	Test	Train	Test	Train	Test	Train	Test
Car	213	66	–	–	707	164	920	230
Camera	541	112	–	–	1197	322	1738	434
Notebook	168	35	–	–	328	88	496	123
Phone	667	156	–	–	1316	341	1983	497
14Lap	870	128	463	169	994	339	2327	636
14Res	807	196	631	196	2164	727	3602	1119
15Res	279	204	36	37	956	349	1271	590
16Res	485	132	72	31	1308	479	1865	642
Twitter	1560	173	3126	345	1561	173	6247	691

Table 3. Hyper-parameter settings in the model.

Parameters	Setting
BERT hidden dimension	768
Dropout rate in BERT	0.1
Learning rate	2e-5
Batch size	16
Training epochs	10
Dropout rate in our model	0.5
Max padding length	80
Optimizer	Adam
Regularization parameter	1×10^{-5}
φ	3
ssh	0.6
α	0.5

4.3. Compared methods

To provide a comprehensive analysis and evaluation of our model, we compared it with several baseline and state-of-the-art models on the ATE and ASC tasks and conducted several ablation experiments for our overall model.

4.3.1. Baseline models

● **ATAE-LSTM** [53] is a neural network model based on attention mechanism and LSTM. By focusing on the aspect target word information from all aspects, it improves the classification effect of fine-grained sentiment analysis tasks.
● **ATSM-S** [47] combines the target-specific memory network with the attention mechanism, using a set of memory units to store the information of the target word and update it based on other words in the context.
● **Sent-Comp** [48] solves the problem of data sparsity by compressing sentences, allowing the model to automatically learn the pragmatic and representative parts of the input data.
● **MenNet** [13] uses multiple memory networks to store the contribution of each word in the context of sentiment polarity classification, combining attention mechanism and word positioning.
● **IAN** [54] uses two LSTMs to introduce an interactive attention mechanism, which can better identify words related to a specific aspect target in the context of a sentence.
● **ASCNN** [55] designs a special CNN structure to capture the contribution of each fragment in the sentence to each aspect target. The input is a sentence, and the output is the sentiment score of each aspect target in the sentence.

4.3.2. State-of-the-art Models

● **BERT-BASE** [19] is a basic version of the pre-trained BERT model released by Google AI Language, which can support the execution of many tasks.

● **BERT-SPC** [43] is a pre-trained BERT model applied to the sentence pair classification task. It differs from BERT-BASE in that the input data format becomes: "[CLS] + sentence + [SEP] + aspect target + [SEP]".

● **SPRN** [56] first obtains the global contextual information and aspect target information of the sentence through the attention mechanism and proposes dual gated multichannel convolution (DGMCC) and dual refinement gate (DRG) to enhance the interaction of sentiment elements between the contexts.

● **MCRF-SA** [57] presents the opinion span of a specific aspect target, which is modeled using multiple CRFs based on that span in combination with a positional decay function.

● **MAN-BERT** [58] uses BERT to replace the transformer encoder in the MAN model.

● **LCF-ATEPC** [44] uses a location mask mechanism to focus sentiment elements on a local context and fuse that local context features with the global context features.

4.3.3. Ablations

● w/o CSE ablates the BERT-based aspect target complex semantic enhancement model part in our model.

● w/o ARE ablated the aspect recognition enhancement mechanism part in our model, which is equivalent to having only BERT-BASE [19] and BERT-SPC [43] models.

● w/o ASD ablates the adaptive semantic distance component in our model.

4.4. Overall performance analysis

For fine-grained sentiment analysis, it is essential to consider not only the APC subtask but also the ATE subtask to ensure the completeness of the process [59,60]. Therefore, the design of a multi-task model is crucial. From an experimental perspective, models based on pre-trained BERT models such as SPRN [56], MCRF-SA [57], MAN-BERT [58], LCF-ATEPC [44], etc. that aggregate aspect target semantic information from different angles have better performance compared to LSTM-based or memory network-based models such as MenNet [13] and IAN [54].

From the results in Table 4, we can draw three conclusions. First, LCF-ATEPC [44], which integrates subtle global and local sentiment information, performs better than BERT-BASE [19] in both tasks, demonstrating the effectiveness of our BERT-ATSE model in multi-level sentiment information fusion. Second, our BERT-ATSE model shows a noticeable improvement in the ATE task after incorporating the complex aspect target semantic enhancement mechanism. Finally, in order to analyze the underwhelming performance on the Camera and Phone datasets, we carefully examined the characteristics of the four Chinese datasets. In comparison to the Car and Notebook datasets, it is undeniable that the Camera and Phone datasets have a larger volume of data. However, the sentences in these datasets are relatively short and have a simpler sentence structure. Additionally, there is a repeated occurrence of the same aspect targets in the corpus. This leads to the model learning relatively monotonous semantics. Since our model primarily serves for sentiment analysis of complex semantics, the experimental improvement on these datasets may be insignificant.

From the results in Table 5, we can draw three conclusions. First, the model LCF-ATEPC [44] with the BERT-SPC [43] input format can significantly improve the identification performance of fine-grained sentiment polarity. Second, our proposed model BERT-ATSE outperforms LCF-ATEPC [44] in Acc and M-F1 values on both tasks in these three datasets, indicating that enhancing the complexity of the corpus and identifying important local semantics in different sentences are essential. Finally,

observing the two tasks in Tables 4 and 5, the BERT-ATSE model performs well in identifying aspect targets and correctly analyzing and judging their sentiment polarities on Chinese and English datasets.

Table 4. Experimental model results (%) on the four Chinese datasets. "-" indicates that this one result is not available. $MF_{1_{ate}}$ indicates the Macro-average F1 value of the aspect target recognition task and Acc_{apc} and $MF_{1_{ate}}$ indicate the Acc and M-F1 value of sentiment polarity classification, respectively.

Model	Car			Camera			Notebook			Phone		
	$MF_{1_{ate}}$	Acc_{apc}	$MF_{1_{apc}}$	$MF_{1_{ate}}$	Acc_{apc}	$MF_{1_{apc}}$	$MF_{1_{ate}}$	Acc_{apc}	$MF_{1_{apc}}$	$MF_{1_{ate}}$	Acc_{apc}	$MF_{1_{apc}}$
Sent-Comp	–	74.38	64.85	–	79.06	67.6	–	83.21	73.72	–	79.65	70.91
ATAE-LSTM	–	81.90	76.88	–	85.54	84.09	–	83.47	82.14	–	85.77	83.87
ATSM-S	–	82.94	64.18	–	82.88	72.50	–	75.59	60.09	–	84.86	75.35
BERT-BASE	86.90	98.26	97.84	86.13	**97.47**	**96.72**	84.62	94.31	93.38	92.10	97.18	96.73
LCF-ATEPC	86.64	97.39	96.72	87.9	96.78	95.86	89.16	94.31	93.29	**92.55**	97.38	**96.96**
BERT-ATSE	**87.24**	**98.27**	**97.86**	**88.32**	96.63	95.79	**90.12**	**95.12**	**93.51**	92.32	**97.51**	**96.96**

Table 5. Experimental results of the two tasks on the three English datasets.

Model	14Laptop			14Restaurant			Twitter		
	$MF_{1_{ate}}$	Acc_{apc}	$MF_{1_{apc}}$	$MF_{1_{ate}}$	Acc_{apc}	$MF_{1_{apc}}$	$MF_{1_{ate}}$	Acc_{apc}	$MF_{1_{apc}}$
BERT-BASE	82.57	79.40	75.24	88.60	82.66	74.13	76.27	73.02	71.43
LCF-ATEPC	82.06	80.03	76.60	88.49	86.06	80.22	95.45	74.23	73.06
BERT-ATSE	**83.50**	**81.23**	**78.57**	**89.51**	**86.79**	**81.41**	**96.73**	**75.87**	**75.02**
BERT-ATSE-MEAN	83.31	80.90	78.23	89.08	86.39	80.98	96.57	75.44	74.73
BERT-ATSE-STD	0.18	0.21	0.32	0.21	0.34	0.27	0.19	0.24	0.31

Our experimental data consists of the best results obtained from 10 different tests. The average values and standard deviations of the 10 test results are presented in Table 5, using the 14Laptop, 14Restaurant and Twitter datasets as examples. The table shows that although the average values are slightly lower than the best results, their standard deviations are small. This indicates that our model exhibits strong stability and consistency.

4.5. Performance analysis of ASC tasks

From the results in Table 6, we can draw several conclusions. First, models with more interaction attention mechanisms, such as MAN-BERT [58], MCRF-SA [57] and BERT-BASE [19], perform similarly on classification tasks. This is mainly because BERT already includes multiple attention heads, making excessive interaction attention redundant. Second, methods that integrate other complex neural network models with BERT, including SPRN [56], perform better than previous methods, demonstrating the effectiveness of this combination approach. Third, when comparing the single-task training of model BERT-ATSE on the APC task to its previous multi-task joint training, we observe a slight decrease in Acc and M-F1 scores. This proves the complementary nature of multi-task joint training, where training for the ATE task and the APC task can promote the learning of model parameters for both tasks. Finally, even in single-task training, the BERT-ATSE model shows slightly lower Acc on the 14Restaurant and 15Restaurant datasets compared to other models. This can be attributed to the presence of a large number of informal and short expressions, as well as the prevalence of ironic sentence patterns in the Restaurant dataset. As a result, the improvement of our model on this dataset is limited. However, other datasets' Acc and M-F1 scores still demonstrate excellent performance, indicating the model's adaptability and robustness.

4.6. Ablation experiments analysis

After conducting ablation experiments on our model, Table 7 provides several conclusions. First, our model BERT-ATSE outperforms other ablated models in terms of Acc and M-F1 values across multiple tasks, demonstrating the reliability of our complex aspect target semantic enhancement. Second, the w/o ARE model performs better in the APC task than the w/o CSE model and the w/o ASD model, indicating that our complex semantic enhancement mechanism and adaptive semantic distance mechanism can have a more significant impact on the model's ability to understand and judge complex sentiment. Third, the w/o ASD model performs better in the ATE task than the w/o CSE model and the w/o ARE model, demonstrating the importance of our complex corpus augmentation mechanism and aspect recognition enhancement mechanism for improving aspect target recognition. Fourth, the w/o CSE model performs better in the APC task in terms of Acc and M-F1 values than the w/o ASD model, highlighting the significant contribution of our adaptive semantic distance mechanism to global and local sentiment semantic understanding and judgment.

4.7. Analysis of the threshold

We tested the sensitivity of the semantic similarity threshold on the camera and 14Laptop datasets and still used BERT-BASE [19] as the underlying structure. Figure 2 shows the training results on these two datasets.

Table 6. Experimental model results (%) for ASC task trained individually on 5 English datasets.

Model	14Laptop		14Restaurant		15Restaurant		16Restaurant		Twitter	
	Acc_{apc}	$MF_{1_{apc}}$	Acc_{apc}	$MF_{1_{apc}}$	Acc_{apc}	$MF_{1_{apc}}$	Acc_{apc}	$MF_{1_{apc}}$	Acc_{apc}	$MF_{1_{apc}}$
MenNet	70.64	65.71	79.61	69.64	77.31	58.28	85.44	65.99	69.65	67.68
IAN	72.05	67.38	79.26	70.09	78.54	52.65	84.74	55.21	71.82	69.11
ASCNN	72.62	66.72	81.73	73.10	78.48	58.90	87.39	64.56	–	–
SPRN	79.31	76.61	85.03	76.97	85.30	–	89.40	–	75.70	73.50
BERT-BASE	79.4	75.24	82.66	74.13	84.54	65.24	88.24	71.18	73.02	71.43
BERT-SPC	78.99	75.03	84.46	76.98	85.91	67.85	89.94	78.23	73.12	71.57
MCRF-SA	77.64	74.23	82.86	73.78	80.82	61.59	89.51	75.92	–	–
MAN-BERT	78.68	75.03	82.05	69.78	85.04	64.98	88.61	74.26	73.99	71.99
LCF-APC	80.50	77.77	**86.15**	80.76	**86.28**	67.66	89.70	78.11	74.24	73.06
BERT-ATSE	**80.87**	**78.04**	86.13	**80.77**	86.23	**68.52**	**90.00**	**78.37**	**74.63**	**74.45**

Table 7. Experimental results of the network structure after ablation.

Model	14Laptop			14Restaurant			Twitter		
	$MF_{1_{ate}}$	Acc_{apc}	$MF_{1_{apc}}$	$MF_{1_{ate}}$	Acc_{apc}	$MF_{1_{apc}}$	$MF_{1_{ate}}$	Acc_{apc}	$MF_{1_{apc}}$
BERT-ATSE	**83.50**	**81.23**	**78.57**	**89.51**	**86.79**	**81.41**	**96.73**	**75.87**	**75.02**
w/o CSE	82.89	80.57	77.86	88.56	85.95	80.69	95.89	74.58	74.01
w/o ARE	82.54	81.03	78.28	87.95	86.24	81.07	95.34	75.11	74.76
w/o ASD	83.13	80.36	76.84	88.56	84.95	80.19	96.21	74.14	73.83

The results presented in Figure 2(a) show that the BERT-ATSE model performs well in terms of Acc and M-F1 score for the APC task on the Camera dataset when the semantic similarity threshold (ssh) is between 0.7 and 0.9. For the ATE task, the best results are obtained when ssh is between 0.7 and 0.8. As ssh increases, the model tends to concatenate sentences with similar semantics to enrich the corpus, and this threshold is more sensitive for the ATE task in this dataset.

The results presented in Figure 2(b) demonstrate that for the APC task of the Laptop dataset, the optimal ssh threshold to achieve the highest Acc and M-F1 score is between 0.5 and 0.6, while for the ATE task, it is at 0.7. Notably, both tasks in this dataset are highly sensitive to changes in the ssh threshold, with the ATE task showing a greater increase in the M-F1 score as ssh increases.
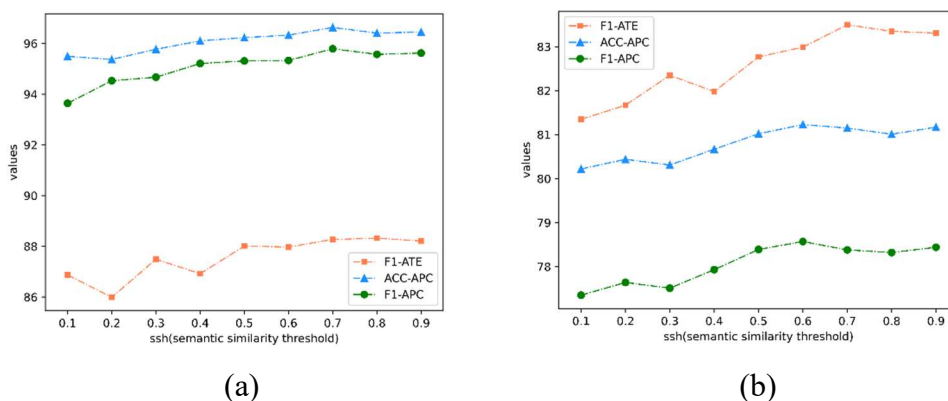
Figure 2. The performance of different semantic similarity thresholds on the Camera and 14Laptop datasets, respectively.

Figure 3. The left sentence illustrates the situation where a sentence contains a single aspect target word, while the right one depicts the scenario where a sentence contains multiple aspect target words. "meal", "lunch", "worth" are the aspect targets, and the dashed box indicates the defined range of concentration of attention in the model.

4.8. Case study

Figure 3 illustrates the attention scores of the best BERT-ATSE model. For the given two input sentences, regarding the first sentence that contains a single aspect target word, BERT-ATSE assigns the aspect term "meal" with the correct negative polarity. It can be observed that the corresponding sentiment terms "terribly" and "thirsty" receive significantly high attention score weights, indicating that they are given greater emphasis in terms of semantic attention. As for the second sentence that contains multiple aspect target words, the two aspect targets, "lunch" and "wait", along with their

corresponding sentiment terms "few times" and "worth", demonstrate that our model, after incorporating ASD, effectively identifies the specific words that each aspect target should pay more attention to, assigning them higher weight values. This results in the correct allocation of neutral polarity and positive polarity to the aspect targets. It can be observed that the length of the sentence does not significantly affect the accurate identification of sentiment terms related to the aspect target. This is because the ASD mechanism intelligently determines the range of redundant information suppression based on the current sentence and aspect target word lengths, playing a crucial role in the model.

5. Conclusions

In this paper, we address the contradiction between the need for more fine-grained sentiment analysis in the ATSC task and the lack of rich aspect target semantics in the available corpora. To tackle this issue, we propose a BERT-based multi-sense learning model that enhances aspect target semantics for both aspect-level sentiment analysis (ATE) and aspect polarity classification (APC) tasks. We use a BERT-based aspect target complex semantic enhancement model to enrich multiple existing training datasets, enabling the model to achieve a higher level of granularity in sentiment analysis. To enhance the robustness of aspect target recognition in ATE tasks, we propose an aspect recognition enhancement mechanism combined with a CRF model to improve the perception of aspect target recognition. Furthermore, we use an adaptive global-local context mechanism to obtain sentiment semantics with substantial overall performance on different datasets after aspect target complex semantic enhancement. Experiments and analysis demonstrate that our model BERT-ATSE can quickly adapt to ATSC tasks and has effectiveness and stability.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

Conflict of interest

The authors declare there is no conflict of interest.

References

1. B. Pang, L. Lee, Opinion mining and sentiment analysis, *Trends Inf. Retr.*, **2** (2008), 1–135. https://doi.org/10.1561/1500000011
2. G. Vinodhini, R. Chandrasekaran, Sentiment analysis and opinion mining: a survey, *Int. J.*, **2** (2012), 282–292. https://doi.org/10.1016/j.nlp.2022.100003

3. M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, S. Manandhar, SemEval-2014 Task 4: Aspect based sentiment analysis, in *Association for Computational Linguistics*, (2014), 27–35 . https://doi.org/10.3115/v1/S14-2004

4. M. Pontiki, D. Galanis, H. Papageorgiou, S. Manandhar, I. Androutsopoulos, Semeval-2015 task 12: Aspect based sentiment analysis, in *Association for Computational Linguistics*, (2015), 486–495. https://doi.org/ 10.18653/v1/S15-2082

5. M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. AL-Smadi, et al. Semeval-2016 task 5: Aspect based sentiment analysis, in *Association for Computational Linguistics*, (2016), 19–30. https://doi.org/ 10.18653/v1/S16-1002

6. W. Zhang, X. Li, Y. Deng, L. Bing, W. Lam, A survey on aspect-based sentiment analysis: Tasks, methods, and challenges, *IEEE Trans. Knowl. Data Eng.*, 2022. https://doi.org/10.1109/TKDE.2022.3230975

7. D. Tang, B. Qin, X. Feng, T. Liu, Effective LSTMs for target-dependent sentiment classification, preprint, arXiv:151201100.

8. M. Yang, W. Tu, J. Wang, F. Xu, X. Chen, Attention based LSTM for target dependent sentiment classification, in *Proceedings of the AAAI conference on artificial intelligence*, 2017. https://doi.org/ 10.1609/aaai.v31i1.11061

9. Q. Liu, Y. Huang, Q. Yang, H. Peng, J. Wang, An attention-aware long short-term memory-like spiking neural model for sentiment analysis, *Int. J. Neural Syst.*, (2023), 2350037–2350037. https://doi.org/ 10.1142/s0129065723500375

10. Y. Huang, Q. Liu, H. Peng, J. Wang, Q. Yang, D. Orellana-Martín, Sentiment classification using bidirectional LSTM-SNP model and attention mechanism, *Expert Syst. Appl.*, **221** (2023), 119730. https://doi.org/10.1016/j.eswa.2023.119730

11. Y. Huang, H. Peng, Q. Liu, Q. Yang, J. Wang, D. Orellana-Martín, et al., Attention-enabled gated spiking neural P model for aspect-level sentiment classification, *Neural Network*, **157** (2023), 437–443. https://doi.org/10.1016/j.neunet.2022.11.006

12. Y. Kim, Convolutional neural networks for sentence classification, preprint, arXiv:14085882.

13. D. Tang, B. Qin, T. Liu, Aspect level sentiment classification with deep memory network, preprint, arXiv:160508900.

14. P. Lin, M. Yang, J. Lai. Deep mask memory network with semantic dependency and context moment for aspect level sentiment classification, in *IJCAI*, (2019), 5088–5094. https://doi.org/10.24963/ijcai.2019/707

15. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., Attention is all you need, in *Advances in Neural Information Processing Systems*, **30** (2017). https://doi.org/10.48550/arXiv.1706.03762

16. Z.-Y. Dou, Capturing user and product information for document level sentiment analysis with deep memory network, in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017. https://doi.org/10.18653/v1/D17-1054

17. K. Chakraborty, S. Bhattacharyya, R. Bag, A survey of sentiment analysis from social media data, *IEEE Trans. Comput. Soc. Syst.*, **7** (2020), 450–464. https://doi.org/ 10.1109/TCSS.2019.2956957

18. X. Zhu, Y. Zhu, L. Zhang, Y. Chen, A BERT-based multi-semantic learning model with aspect-aware enhancement for aspect polarity classification, *Appl. Intell.*, **53** (2023), 4609–4623. https://doi.org/ 10.1007/s10489-022-03702-1

19. J. Devlin, M. W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, preprint, arXiv:181004805.

20. N. Reimers, I. Gurevych, Sentence-bert: Sentence embeddings using siamese bert-networks, preprint, arXiv:190810084.

21. L. Breiman, J. Friedman, C. J. Stone, R. A. Olshen, Classification and Regression Trees (CART), *Biometrics*, **1984** (1984). https://doi.org/10.2307/2530946

22. N. S. Altman, An introduction to kernel and nearest-neighbor nonparametric regression, *Am. Stat.*, **46** (1992), 175–185. https://doi.org/10.1080/00031305.1992.10475879

23. I. Rish, An empirical study of the naive Bayes classifier, in *IJCAI 2001 workshop on empirical methods in artificial intelligence*, (2001), 41–46. https://doi.org/10.1109/CSCI46756.2018.00065

24. D. W. Hosmer Jr, S. Lemeshow, R. X. Sturdivant, *Applied Logistic Regression*, John Wiley & Sons, 2013. https://doi.org/10.1002/9781118548387

25. C. Cortes, V. Vapnik, Support-vector networks, *Mach. Learn.*, **20** (1995), 273–297.

26. L. Breiman, Random forests, *Mach. Learn.*, **45** (2001), 5–32. https://doi.org/10.1023/A:1022627411411

27. N. S. Joshi, S. A. Itkat, A survey on feature level sentiment analysis, *Int. J. Comput. Sci. Inf. Technol.*, **5** (2014), 5422–5425.

28. E. Cambria, B. White, Jumping NLP curves: A review of natural language processing research, *IEEE Comput. Intell. Mag.*, **9** (2014), 48–57. https://doi.org/10.1109/MCI.2014.2307227

29. B. Zhang, X. Fu, C. Luo, Y. Ye, X. Li, L. Jing, Cross-domain aspect-based sentiment classification by exploiting domain-invariant semantic-primary feature, *IEEE Trans. Affect. Comput.*, **2023** (2023), forthcoming. https://doi.org/10.1109/TAFFC.2023.3239540

30. H. Huang, B. Zhang, L. Jing, X. Fu, X. Chen, J. Shi, Logic tensor network with massive learned knowledge for aspect-based sentiment analysis, *Knowl. Based Syst.*, **257** (2022), 109943. https://doi.org/10.1016/j.knosys.2022.109943

31. X. Mei, Y. Zhou, C. Zhu, M. Wu, M. Li, S. Pan, A disentangled linguistic graph model for explainable aspect-based sentiment analysis, *Knowl. Based Syst*, **260** (2023), 110150. https://doi.org/10.1016/j.knosys.2022.110150

32. B. Zhang, X. Huang, Z. Huang, H. Huang, B. Zhang, X. Fu, et al., Sentiment interpretable logic tensor network for aspect-term sentiment analysis, in *Proceedings of the 29th International Conference on Computational Linguistics*, (2022), 6705–6714.

33. B. Xu, X. Wang, B. Yang, Z. Kang, Target embedding and position attention with lstm for aspect based sentiment analysis, in *Proceedings of the 2020 5th International Conference on Mathematics and Artificial Intelligence*, (2020), 93–97. https://doi.org/10.1145/3395260.3395280

34. Y. Ma, H. Peng, E. Cambria, Targeted aspect-based sentiment analysis via embedding commonsense knowledge into an attentive LSTM, in *Proceedings of the AAAI conference on artificial intelligence*, (2018). https://doi.org/10.1609/aaai.v32i1.12048

35. L. Bao, P. Lambert, T. Badia, Attention and lexicon regularized LSTM for aspect-based sentiment analysis, in *Proceedings of the 57th annual meeting of the association for computational linguistics: student research workshop*, (2019), 253–259. https://doi.org/10.18653/v1/P19-2035

36. Y. Xing, C. Xiao, Y. Wu, Z. Ding, A convolutional neural network for aspect-level sentiment classification, *Int. J. Pattern Recognit. Artif Intell.*, **33** (2019), 1959046. https://doi.org/10.18653/v1/2021.textgraphs-1.8

37. X. Wang, F. Li, Z. Zhang, G. Xu, J. Zhang, X. Sun, A unified position-aware convolutional neural network for aspect based sentiment analysis, *Neurocomputing*, **450** (2021), 91–103. https://doi.org/10.1016/j.neucom.2021.03.092

38. C. Gan, L. Wang, Z. Zhang, Z. Wang, Sparse attention based separable dilated convolutional neural network for targeted sentiment analysis, *Knowl. Based Syst.*, **188** (2020), 104827. https://doi.org/10.1016/j.knosys.2019.06.035

39. N. Zhao, H. Gao, X. Wen, H. Li, Combination of convolutional neural network and gated recurrent unit for aspect-based sentiment analysis, *IEEE Access*, **9** (2021), 15561–15569. https://doi.org/10.1109/ACCESS.2021.3052937

40. Y. Tay, L. A. Tuan, S. C. Hui, Dyadic memory networks for aspect-based sentiment analysis, in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, (2017), 107–116. https://doi.org/10.1145/3132847.3132936

41. Y. Chen, T. Zhuang, K. Guo, Memory network with hierarchical multi-head attention for aspect-based sentiment analysis, *Appl. Intell.*, **51** (2021), 4287–4304. https://doi.org/10.1007/s10489-020-02069-5

42. Y. Zhang, B. Xu, T. Zhao, Convolutional multi-head self-attention on memory for aspect sentiment classification, *IEEE-CAA J. Automatica Sin.*, **7** (2020), 1038–1044. https://doi.org/10.1109/JAS.2020.1003243

43. Y. Song, J. Wang, T. Jiang, Z. Liu, Y. Rao, Attentional encoder network for targeted sentiment classification, preprint, arXiv:190209314.

44. H. Yang, B. Zeng, J. Yang, Y. Song, R. Xu, A multi-task learning model for chinese-oriented aspect polarity classification and aspect term extraction, *Neurocomputing*, **419** (2021), 344–356. https://doi.org/10.1016/j.neucom.2020.08.001

45. A. Karimi, L. Rossi, A. Prati, Improving bert performance for aspect-based sentiment analysis, preprint, arXiv:201011731.

46. A. Karimi, L. Rossi, A. Prati, Adversarial training for aspect-based sentiment analysis with bert, in *2020 25th International conference on pattern recognition (ICPR)*, (2021), 8797–8803. https://doi.org/10.1109/ICPR48806.2021.9412167

47. H. Peng, Y. Ma, Y. Li, E. Cambria, Learning multi-grained aspect target sequence for Chinese sentiment analysis, *Knowl. Based Syst.*, **148** (2018), 167–176.

48. W. Che, Y. Zhao, H. Guo, Z. Su, T. Liu, Sentence compression for aspect-based sentiment analysis, *IEEE-ACM Trans. Audio Speech Lang.*, **23** (2015), 2111–2124.

49. L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, K. Xu, Adaptive recursive neural network for target-dependent twitter sentiment classification, in *Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 2: Short papers)*, (2014), 49–54.

50. B. Wang, W. Lu, Learning latent opinions for aspect-level sentiment classification, in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2018.

51. H. T. Nguyen, M. Le Nguyen, Effective attention networks for aspect-level sentiment classification, in *2018 10th International Conference on Knowledge and Systems Engineering (KSE)*, (2018), 25–30. https://doi.org/10.1109/KSE.2018.8573324

52. D. P. Kingma, J. Ba, Adam: A method for stochastic optimization, preprint, arXiv:14126980.

53. Y. Wang, M. Huang, X. Zhu, L. Zhao, Attention-based LSTM for aspect-level sentiment classification, in *Proceedings of the 2016 conference on empirical methods in natural language processing*, (2016), 606–615. https://doi.org/10.18653/v1/D16-1058

54. D. Ma, S. Li, X. Zhang, H. Wang, Interactive attention networks for aspect-level sentiment classification, preprint, arXiv:170900893.

55. H. Peng, L. Xu, L. Bing, F. Huang, W. Lu, L. Si, Knowing what, how and why: A near complete solution for aspect-based sentiment analysis, in *Proceedings of the AAAI conference on artificial intelligence*, (2020), 8600–8607. https://doi.org/10.1609/aaai.v34i05.6383

56. W. Song, Z. Wen, Z. Xiao, S. C. Park, Semantics perception and refinement network for aspect-based sentiment analysis, *Knowl. Based Syst.*, **214** (2021), 106755.

57. L. Xu, L. Bing, W. Lu, F. Huang, Aspect sentiment classification with aspect-specific opinion spans, in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, (2020), 3561–3567. https://doi.org/10.18653/v1/2020.emnlp-main.288

58. Q. Xu, L. Zhu, T. Dai, C. Yan, Aspect-based sentiment classification with multi-attention network, *Neurocomputing*, **388** (2020), 135–143. https://doi.org/10.1016/j.neucom.2020.01.024

59. B. Huang, J. Zhang, J. Ju, R. Guo, H. Fujita, J. Liu, CRF-GCN: An effective syntactic dependency model for aspect-level sentiment analysis, *Knowl. Based Syst.*, **260** (2023), 110125. https://doi.org/10.1016/j.knosys.2022.110125

60. B. Huang, R. Guo, Y. Zhu, Z. Fang, G. Zeng, J. Liu, et al., Aspect-level sentiment analysis with aspect-specific context position information, *Knowl. Based Syst.*, **243** (2022), 108473. https://doi.org/10.1016/j.knosys.2022.108473